

# Kinetic Stability and Crystal Structure of the Viral Capsid Protein SHP

Patrik Forrer<sup>1</sup>†, Changsoo Chang<sup>2</sup>†, David Ott<sup>1</sup>, Alexander Wlodawer<sup>2</sup>  
and Andreas Plückthun<sup>1\*</sup>

<sup>1</sup>Biochemisches Institut  
Universität Zürich  
Winterthurerstrasse 190  
CH-8057 Zürich, Switzerland

<sup>2</sup>Macromolecular  
Crystallography Laboratory  
National Cancer Institute at  
Frederick, Frederick, MD 21702  
USA

SHP, the capsid-stabilizing protein of lambdaoid phage 21, is highly resistant against denaturant-induced unfolding. We demonstrate that this high functional stability of SHP is due to a high kinetic stability with a half-life for unfolding of 25 days at zero denaturant, while the thermodynamic stability is not unusually high. Unfolding experiments demonstrated that the trimeric state (also observed in crystals and present on the phage capsid) of SHP is kinetically stable in solution, while the monomer intermediate unfolds very rapidly. We also determined the crystal structure of trimeric SHP at 1.5 Å resolution, which was compared to that of its functional homolog gpD. This explains how a tight network of H-bonds rigidifies crucial interpenetrating residues, leading to the observed extremely slow trimer dissociation or denaturation. Taken as a whole, our results provide molecular-level insights into natural strategies to achieve kinetic stability by taking advantage of protein oligomerization. Kinetic stability may be especially needed in phage capsids to allow survival in harsh environments.

© 2004 Elsevier Ltd. All rights reserved.

**Keywords:** bacteriophage; kinetic stability; protein oligomerization; protein structure; virus stability

\*Corresponding author

## Introduction

Protein stability involves both thermodynamic and kinetic aspects.<sup>1–3</sup> The thermodynamic stability of proteins can be marginal,<sup>4,5</sup> but even a high thermodynamic stability of proteins does not prevent their unfolding *per se*. Only the ratio between the unfolded and folded states of a protein at equilibrium is described by thermodynamics, but not their rate of interconversion. The kinetic stability of a protein is the result of an energy barrier that exists between the folded and unfolded states. This barrier may be very high, making the unfolding reaction very slow on the biologically relevant time scale, and thereby stabilizing the folded state of the protein even if the thermodynamic stability is low. More and more proteins or protein complexes are found to possess high kinetic stability, including bacterial luciferase,<sup>6</sup> influenza virus hemagglutinin,<sup>7</sup> trans-thyretin,<sup>8</sup> recombinant mouse prion protein,<sup>9</sup>

cyanovirin-N,<sup>10</sup> and the synaptic SNARE complex.<sup>11</sup> In extreme cases, kinetic stability can prevent the unfolding of the native state on a biologically relevant time-scale, even when the native state is thermodynamically unstable, as reported for the bacterial alpha-lytic protease.<sup>12,13</sup> All these recent findings demonstrate that kinetic stability is an important flavor of protein stability and is indispensable for many biological functions.

Our preliminary work on SHP, a major capsid protein from the lambdaoid phage 21,<sup>14</sup> also indicated that it possesses a high kinetic stability. This attracted our interest, as some viruses are known to possess high kinetic stability,<sup>15</sup> but the underlying molecular mechanisms are not well understood. Once virus particles are released from their hosts, they have to resist the harsh conditions of the extracellular environment until they find a new host to infect. Thus, their assembly has to be “quasi-irreversible” in order for the virus not to fall apart before infection. Such a high stability is equivalent to very high activation energy for unfolding, but this energy barrier does not necessarily have to correspond to an unusually high thermodynamic stability. We were thus interested to learn more about the observed kinetic stability of SHP by analyzing its unfolding pathway and its three-dimensional

† P.F. & C.C. contributed equally to this work.

Abbreviations used: EM, electron microscopy; r.m.s., root-mean-square; GdmCl, guanidinium chloride; RT, room temperature.

E-mail address of the corresponding author: [plueckthun@bioc.unizh.ch](mailto:plueckthun@bioc.unizh.ch)

structure. Our results provide general insights into possible strategies to achieve kinetic stability on a molecular level.

## Results

### Expression and purification of SHP

SHP is the functional homolog of protein D (gpD),<sup>16</sup> the capsid-stabilizing protein of bacteriophage  $\lambda$ . Protein D is added last in the phage assembly and it has been used as a fusion partner for phage  $\lambda$  display.<sup>17</sup> The mature  $\lambda$ -head contains 405–420 copies of gpD arranged into well separated “thimble-shaped” protrusions, interpreted as trimers of gpD, overlying the trigonal sites of the capsid.<sup>18</sup> We previously solved the crystal structure of trimeric gpD and demonstrated its correspondence to the capsid-bound trimer by modeling of the structure into a 15 Å resolution cryo-electron microscopy (EM) density map of the mature  $\lambda$ -capsid.<sup>19</sup> Since SHP is able to functionally replace gpD on the  $\lambda$ -capsid, its mechanism of binding and capsid stabilization should be the same as that of gpD. In common with gpD,<sup>20</sup> with which it shares 50% sequence identity, SHP is very well expressed and highly soluble.<sup>19</sup> gpD and SHP are both small cysteine-free proteins of 109 and 114 amino acid residues in length, respectively (not counting the initiator methionine, which is not present in either mature protein). SHP was expressed and purified to near homogeneity as described.<sup>19</sup>

### The crystal structure of SHP

The structure of SHP was solved by molecular replacement in two crystal forms (Table 1). Monoclinic form I crystals (space group C2) contain a trimer in the asymmetric unit. The quality of these

crystals is relatively poor, resulting in high mosaic spread and a limited extent of diffraction, thus causing difficulty in measuring accurate data. Rhombohedral form II crystals (space group R3) contained only a monomer in the asymmetric unit, although a trimer almost identical to the one seen for form I (r.m.s. deviation of 0.64 Å) was generated by crystallographic symmetry operators  $(-y, x-y, z)$  and  $(-x+y, -x, z)$ . These crystals diffracted to high resolution and the resulting model could be refined quite well, to the final *R*-value of 12.6% and *R*<sub>free</sub>-value of 18.5% (Table 1).

The atomic *B*-factors of the final model for form I are relatively high, thus the resulting electron density map is of comparatively low quality. The mean positional error in atomic coordinates as estimated by the Luzzati plot is 0.29 Å. All non-glycine and non-proline residues of the model lie in either the most favorable region or in the additionally allowed region of the Ramachandran plot. For form II, the electron density map is excellent in all regions. The mean positional error in atomic coordinates as estimated by the Luzzati plot is 0.13 Å. Except for one residue, all non-glycine and non-proline residues of the model are found in either the most favorable region or in the additionally allowed region of the Ramachandran plot. The only residue found in the generously allowed region is Ser93, which has been assigned the very unusual *D*-configuration (see below). Due to the much higher quality of the form II structure, it is utilized here for all analysis and comparisons, unless explicitly mentioned otherwise.

The two structures of SHP are almost identical, except for the N and C termini. The r.m.s. deviation between all C $\alpha$  positions of SHP in crystal forms I and II is 0.59 Å, but when the two termini (residues 12 and 114) are excluded, the r.m.s. deviation is only 0.36 Å. Generally, the overall fold of SHP is the same as that of gpD. Sixty-seven out of 102 residues belong to an irregular secondary structure such as turns or coil. One monomer contains seven  $\beta$ -strands (S1, 23–27; S2, 39–42; S3, 49–51; S4, 62–64; S5, 74–78; S6, 80–84; and S7, 107–111), one  $\alpha$ -helix (residues 95–101), and one  $_3_{10}$  helix (86–88). The strands S1 through S5 create an antiparallel  $\beta$ -sheet with the order S1, S5, S4, S3, and S2, whereas S6 and S7 are parallel. The hydrogen bonds between the strands in the antiparallel  $\beta$ -sheet are quite irregular and their number is low. Strands S1 and S5 are connected by four hydrogen bonds, there are only two hydrogen bonds between strands S5 and S4, none between S4 and S3, and two between strands S2 and S3. For the parallel  $\beta$ -sheet, only two hydrogen bonds link the two five-residue strands. The pattern of hydrogen bonds closely follows the situation found in gpD. Similarly to gpD, the overall packing of SHP is tight and without any internal cavities, creating a dense hydrophobic core. This core is composed of Pro39, Leu40, Trp52, Ala60, Gly62, Ile63, Leu64, Gly81, Phe83, Ala84, Trp 90, and Ala102. The trimer of SHP is shaped as a triangle with a rounded side and closely

**Table 1.** Data collection and refinement statistics

	SHP form I	SHP form II
<b>A. Data collection</b>		
Wavelength (Å)	1.54	0.98
Space group	C2	R3
Unit cell parameters (Å)	$a=100.1$ ; $b=57.7$ ; $c=62.0$ ; $\beta=117.1^\circ$	$a=b=53.9$ ; $c=77.5$
Resolution (Å)	60–2.37	40–1.50
Total number of reflections	238,035	135,607
Unique reflections	12,946	13,405
Completeness (last shell) (%)	99.8 (100.0)	99.8 (98.4)
<i>R</i> <sub>merge</sub> (%)	8.8 (38.5)	8.5 (46.1)
<b>B. Data refinement</b>		
No. of monomers in a.u.	3	1
Resolution range (Å)	45–2.37	40–1.50
Reflections used	12,465	10,706
<i>R</i> <sub>cryst</sub> (%)	21.0	12.6
<i>R</i> <sub>free</sub> (%)	24.3	18.5
No. of protein atoms	2242	758
No. of solvent atoms	387	167
r.m.s. deviations from ideality:		
Bond lengths (Å)	0.007	0.011
Angles (°)/bond angles (Å)	1.4	0.028

**Table 2.** Intersubunit hydrogen bonds

SHP	GpD
Asn101 OD1-N Ser27	–
Asn101 ND2-O Ser27	–
Asn101 ND2-O Gly47	Thr99 OG-O Arg44
Lys79 NZ-O Ala107	Lys76 NZ-O Ala105
Lys79 NZ-OD2 Asp19	Lys76 NZ-OD2 Asp16
Lys79 NZ-OG Ser109	–
Lys56 NZ-OE2 Glu44	–

corresponds to the trimer of gpD. Accordingly, the “bottom” and “top” sides were defined as for gpD.<sup>19</sup> In creating a trimer, 16 residues from each monomer are involved in non-bonding interactions with the other monomers; they are Asp19, His22, Ala24, Ser27, Leu42, Gly47, Lys48, Ala57, Gly58, Val61, Lys79, His97, Ala100, Asn101, Gly105, and Ser109. The residues involved in intersubunit hydrogen bonds are listed in Table 2. No intermolecular hydrogen bonds are made between the main-chains of the monomers, whereas 21 intermolecular hydrogen bonds involve either main-chain and side-chain atoms, or side-chain atoms only. Most of the interactions within the trimer interface are hydrophobic in nature. The surface area buried on trimerization is 1227 Å<sup>2</sup>, which corresponds to 21.6% of the surface of the monomer.

Serine 93 appears to be in a very rare D-configuration in the form II structure of SHP. A search of all structures deposited in the RCSB Protein Data Bank as of February 2001 has shown only two proteins with D-amino acids included in their main-chains. The structure of endothiapepsin<sup>21</sup> (PDB accession code 1EPR) included Glu135 as D-amino acid, while Asp115 was found in a D-configuration in a complex of porcine trypsin with an inhibitor (PDB accession code 1AN1). In the latter case, the D-configuration is explained by the backbone rearrangement *via* deamidation, followed by selection of the conformer during prolonged crystallization.<sup>22</sup> No reports of a similar phenomenon could be found for a serine, so our identification of the D-amino acid, although based on the results of high-resolution refinement, must be taken with some caution. While acid catalyzed dehydration/rehydration with racemization at pH 4.5 in the crystallization buffer and selective crystallization of one form would provide a conceivable mechanism for incorporation of the D-Ser, we have no evidence for this, nor do we suggest a biological function.

### Structural comparison of SHP and gpD

The overall structure of SHP is very similar to that of gpD. When the coordinates of a monomer of SHP were superimposed on the coordinates of gpD (Figure 1(a)), the r.m.s. deviation for 94 C<sup>α</sup> atom pairs in residues 18–111 was 1.09 Å. When the trimers were compared, the r.m.s. deviation for 279 C<sup>α</sup> pairs was 1.4 Å. Significant differences are present in the vicinity of Ala57 (Asp54 in gpD). These residues are located at the mouth of the top cleft, and the r.m.s. deviation of their C<sup>α</sup> positions is 4.61 Å. Ala57 is

located in the loop between strands S3 and S4. This loop (residues 52–59 in SHP; 49–56 in gpD) has more interactions with the neighboring residues from the same subunit than in gpD, making four hydrogen bonds in the former and only one in the latter. In addition, this loop is involved in hydrophobic interactions with seven neighboring residues in SHP, but with only five neighboring residues in gpD. As a result, the mouth of the top cleft in SHP is smaller than that of gpD; the length of the triangular mouth edge is 16 Å in SHP, but 18 Å in gpD.

Another location with significant differences between SHP and gpD is found near Ser93. The r.m.s. deviation of the C<sup>α</sup> position of Ser93 with the corresponding residue of gpD is 3.24 Å and a one-residue gap can be seen in the alignment of the two proteins. The N and C<sup>α</sup> atoms of Glu92 (Glu89 in gpD) superimpose well, but the  $\psi$  angle of this residue is very different (144° in SHP, –34° in gpD); torsion angles of residues 93 and 94 in SHP (residues 90, 91, and 92 in gpD) are also different. For these reasons, the path taken by the main-chain is somewhat different in this region. The residues following Asp95 (93 in gpD) have similar torsion angles although the position of the backbone atoms is also slightly different.

The areas showing considerable variation between SHP and gpD are found at their N and C termini. The N-terminal sequences of these two proteins do not seem to be related and are disordered in both proteins, although to a different extent. The ordered part of SHP starts at residue 12, whereas the first 14 residues are disordered in gpD.<sup>19</sup> The additional visible residues in SHP extend out to the edge of the molecule. Considering intermolecular interactions in the N-terminal region, Pro20 (Pro17 in gpD) makes a hydrophobic interaction with His22 (His19 in gpD) of the adjacent molecules, and Asp19 (Asp16 in gpD) makes ion pair contact with Lys79 (Lys76 in gpD) of the adjacent molecule. These intermolecular interactions are found also in the structure of gpD. The residues additionally defined in SHP do not make any intermolecular interactions, although they make intramolecular interactions with neighboring residues. Nevertheless, they may help to stabilize the orientation of those residues that are involved in interactions near the 3-fold axis. Asn18 makes no interactions with any other residues, and then starting from residue 17 to the N terminus, the direction of the chain is pointing not at an adjacent molecule, but rather away from the 3-fold axis toward the edge of the triangle along the strand S5 of the same molecule. Ala16 makes a main-chain hydrogen bond with Thr82, whereas Ile14 makes such a bond with Ala84. In addition, these N-terminal residues make hydrophobic interactions with Ala21, Leu37, Phe83, and Leu113 (Figure 1(b)).

A comparison of the interface<sup>†</sup> showed that the accessible surface buried upon trimer formation,

<sup>†</sup> <http://www.biochem.ucl.ac.uk/bsm/PP/server>

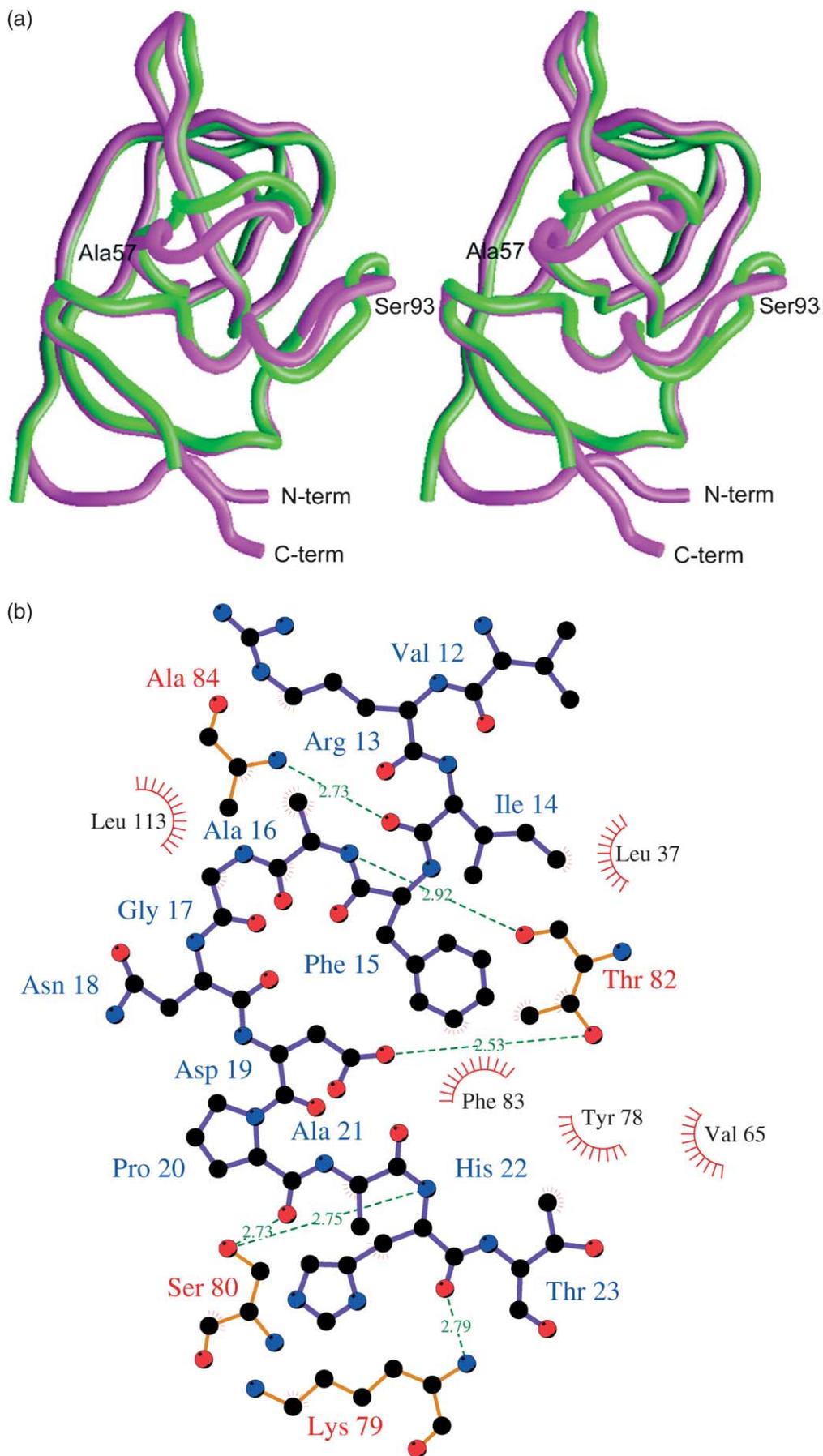
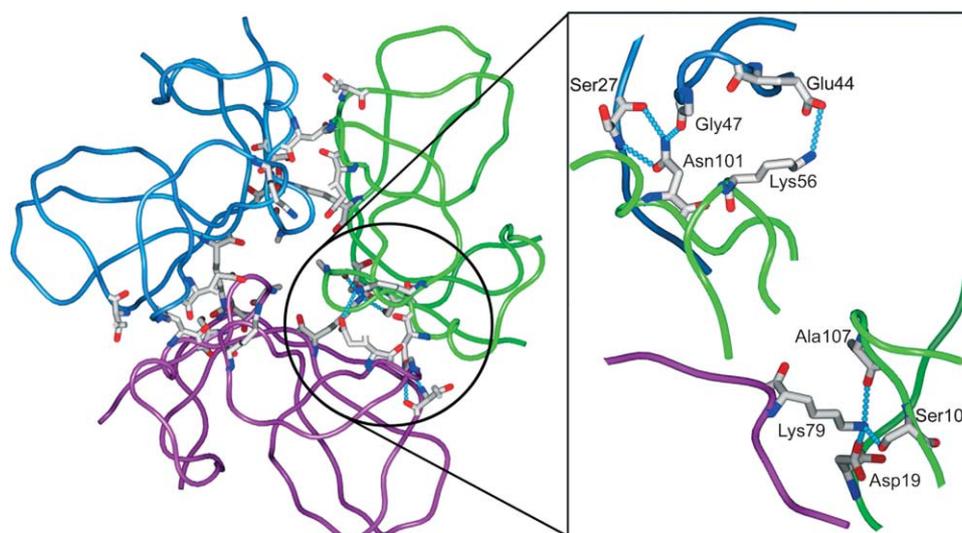


Figure 1 (legend opposite)



**Figure 2.** Inter-subunit hydrogen bonds. The hydrogen bonds between different subunits of SHP are marked in dotted lines. Backbone traces of each subunit in a trimer are colored blue, purple, and green, respectively. The details within one monomer are shown enlarged in an inset. The residues involved in forming inter-subunit hydrogen bonds are shown in stick representation, and they are also listed in Table 2.

assessed with probe size of  $1.4 \text{ \AA}$ , is very similar ( $1220 \text{ \AA}^2$  for each subunit of gpD,  $1227 \text{ \AA}^2$  for SHP), with a very similar gap volume. In the SHP trimer, however, there are a total of 21 intersubunit hydrogen bonds, while only nine are present within the gpD trimer (Table 2). In SHP, three areas are involved in intersubunit H-bonding interactions (Table 2; Figure 2). The side-chain of Asn101 makes three H-bonds, two to the main-chain of Ser27 of the neighboring subunit and one to the carbonyl oxygen of Gly47. Only this last H-bond is found in gpD, between the corresponding positions of Thr99 and the carbonyl oxygen of Arg44. The second area is found around Lys79, which donates three H-bonds in SHP, to the carbonyl oxygen of Ala107, the side-chain of Asp19, and the side-chain of Ser109. While all these residues are conserved in gpD, the last interaction (between Lys76 and Ser107) is of poor geometry and will thus not energetically contribute very much and is not counted here as an H-bond. Finally, a side-chain interaction between the charged Lys56 and Glu44 in SHP has no equivalent in gpD, because both of the loops involved in its creation have a slightly different orientation of the side-chains and a very different sequence.

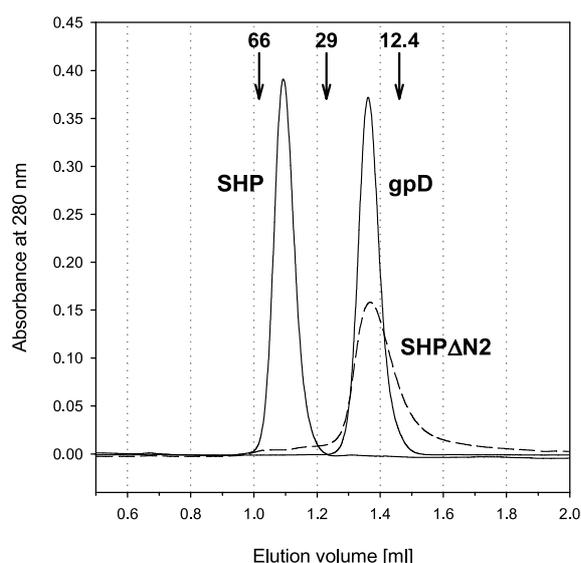
### SHP is a highly stable trimer in solution

SDS-PAGE analysis of the purified gpD and SHP revealed the expected band of 11 kDa for gpD,

whereas two different forms of SHP were detectable by Coomassie staining, one at 11 kDa and one around 32 kDa (data not shown). The unexpected high molecular mass band of SHP completely disappeared only when the sample was extensively heated ( $>20$  minutes) in the SDS loading buffer before being applied to the gel. Since SHP has no cysteine residues, this behavior cannot be due to disulfide formation and thus indicates the existence of a highly stable SHP species that denatured only very slowly upon heating in the SDS loading buffer. Since we considered that the high molecular mass band found for SHP might correspond to the trimeric state analogous to that observed in the crystal structure (Figure 2), we analyzed the proteins by gel filtration (Figure 3). The apparent molecular mass of SHP (50.2 kDa) was almost exactly three times that of gpD (17.6 kDa). This indicated that only SHP, in contrast to gpD, forms a stable trimer in solution, although both are trimers in the crystals and when bound to the phage capsid.<sup>19</sup>

The structures of both SHP (Figure 2) and gpD<sup>19</sup> revealed a highly conserved feature consisting of a small ring (Pro/His-ring) near the 3-fold axis at the bottom side of the trimers. This ring is formed by two aromatic residues from each monomer (Pro21 and His23 in SHP; Pro17 and His19 in gpD). The formation of this ring may play a key role in the trimerization of these proteins. To test this hypothesis, we constructed a His-tagged N-terminal

**Figure 1.** (a) Stereo diagram showing the superposition of SHP and gpD monomers. The backbone trace of SHP is presented in magenta and the trace of gpD in green. The residues showing large positional deviations between SHP and gpD are marked with their names and numbers. The termini of SHP are also marked. (b) Inter-subunit interactions of the N-terminal region (Val12 to Thr23) of SHP, generated using the program LIGPLOT.<sup>47</sup> Interatomic bonds within the N-terminal residues are in purple, whereas other residues are colored brown. Hydrogen bonds are marked as green dotted lines with explicitly given bond lengths. Hydrophobic interactions are shown as semicircles.

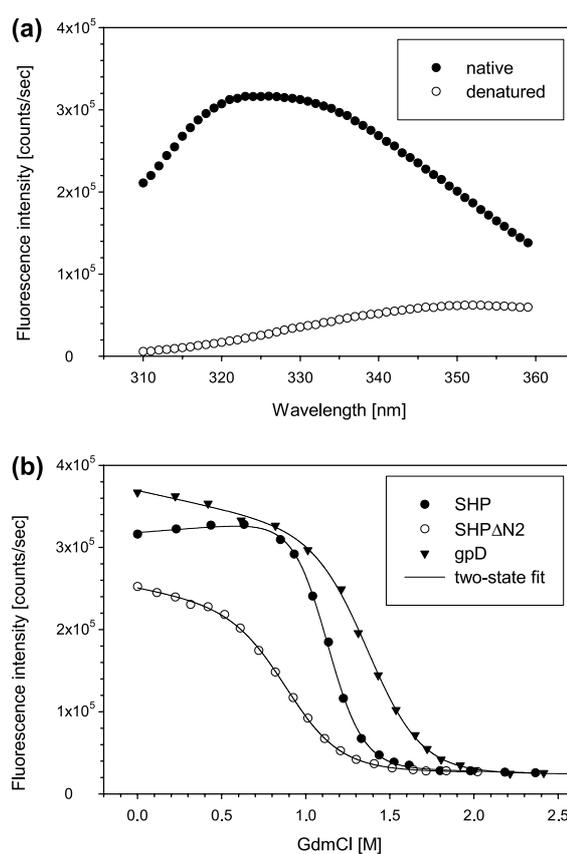


**Figure 3.** Gel filtration elution profiles of SHP, SHP $\Delta$ N2 and gpD. Samples (50  $\mu$ l) were loaded at a monomer concentration of 50  $\mu$ M on a Superdex 75 column and were run at 60  $\mu$ l  $\text{min}^{-1}$  in TBS<sub>150</sub>. Arrows indicate the elution volumes of marker proteins (cytochrome *c*, 12.4 kDa; carbonic anhydrase, 29 kDa; bovine serum albumin, 66 kDa). The apparent molecular masses of SHP (11.8 kDa), SHP $\Delta$ N2 (10.8 kDa) and gpD (11.4 kDa) are 50.2, 16.9 and 17.6 kDa, respectively. Identical results were obtained for concentrations down to at least 0.5  $\mu$ M.

deletion variant of SHP (SHP $\Delta$ N2), where residues Val1 to His23 were deleted. Purified SHP $\Delta$ N2 instantly denatures in SDS loading buffer, in contrast to wild-type SHP (data not shown), and was found to be monomeric in solution (Figure 3). Thus, the behavior of SHP $\Delta$ N2 resembles the solution behavior of gpD, but not that of the full-length SHP. A control construct gpD $\Delta$ N2 (N-terminal deletion variant of gpD with removed Thr1 to His19), which corresponds exactly to SHP $\Delta$ N2, behaved in the same way as wild-type gpD (data not shown). This implies that SHP $\Delta$ N2 is a valid model for monomeric SHP. Taken together, the high stability observed for SHP in solution seems to correlate well with its trimeric state and the presence of the N-terminal amino acid residues up to and including the residues of the Pro/His-ring.

### Equilibrium unfolding of SHP

Equilibrium unfolding curves for SHP, SHP $\Delta$ N2, and gpD were determined using guanidinium chloride (GdmCl)-induced unfolding monitored by Trp fluorescence at monomer concentrations of 0.5  $\mu$ M. Figure 4(a) shows the fluorescence emission spectra of SHP in its native state and after denaturation in 2.4 M GdmCl. SHP contains two Trp residues at positions 52 and 90, both buried in the hydrophobic core of the native molecule. Both residues become solvent exposed upon denaturation, resulting in a red shift of the fluorescence



**Figure 4.** Unfolding measured by fluorescence. (a) Fluorescence emission spectra of native and in 2.4 M GdmCl denatured SHP. Protein excitation was at 295 nm. (b) Equilibrium unfolding curves of SHP, SHP $\Delta$ N2 and gpD, monitored by the change in fluorescence intensity at 325 nm. Data were fitted to a two-state model. The concentration of all proteins was 0.5  $\mu$ M. Denaturant concentrations at the midpoint of transition,  $[D]_{1/2}$ , of 1.1, 0.9 and 1.4 M GdmCl were determined for SHP, SHP $\Delta$ N2 and gpD, respectively.

emission maximum from 325 nm to 351 nm and in a concomitant reduction of the fluorescence intensity (excitation wavelength was 295 nm). Both Trp residues are conserved in gpD. The corresponding emission spectra for gpD, gpD $\Delta$ N2, and SHP $\Delta$ N2 are very similar to those measured for wild-type SHP (data not shown). Figure 4(b) shows the equilibrium unfolding curves for SHP, SHP $\Delta$ N2 and gpD, obtained by following the fluorescence intensity at 325 nm. All three proteins showed cooperative and reversible unfolding. The individual unfolding curves gave midpoints of transition,  $[D]_{1/2}$ , of 1.1, 0.9 and 1.4 M GdmCl for SHP, SHP $\Delta$ N2 and gpD, respectively. However, SHP samples had to be incubated for at least ten days before measurement in order to reach equilibrium, as we observed strong kinetic hysteresis between the renaturation and denaturation transitions. The hysteresis is indicative of a high kinetic barrier present between unfolded and folded species of SHP. By contrast, the samples of gpD and SHP $\Delta$ N2

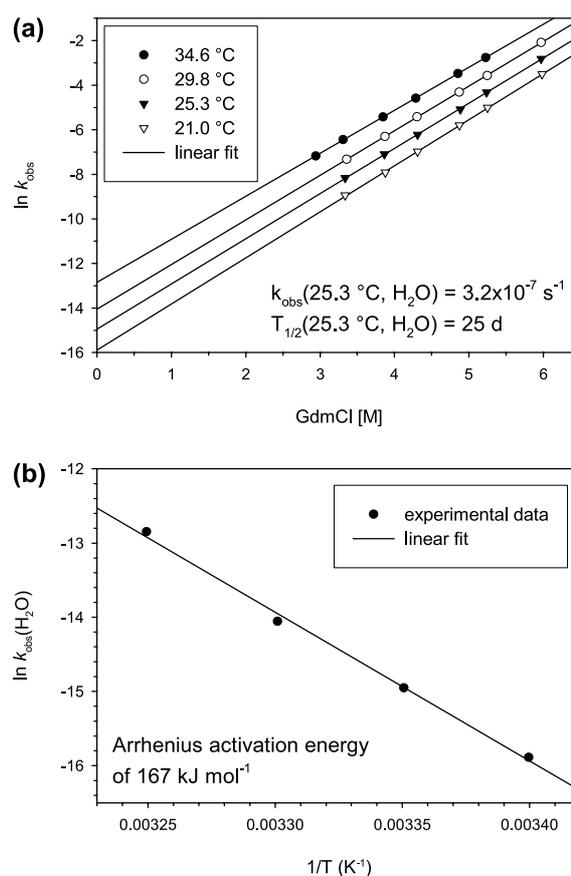
were equilibrated in minutes. The steeper transition in the unfolding curve of SHP, when compared to gpD, is indicative for a higher cooperativity in the unfolding of SHP, as it would be expected for a trimer compared to a monomer.

However, the existence of a stable folded monomeric kinetic intermediate in the folding pathway of SHP (see below) indicated that the equilibrium unfolding of SHP may not follow a simple two-state model of folded trimers and unfolded monomers and that a stable monomer may also be a thermodynamic intermediate; this essentially precludes the extraction of thermodynamic parameters for SHP. Monomeric SHP $\Delta$ N2 and gpD yield values for the free energy of unfolding ( $\Delta G_u$ ) of 13.8 kJ mol<sup>-1</sup> and 21.6 kJ mol<sup>-1</sup>, and  $m$ -values of 15.4 kJ mol<sup>-1</sup> M<sup>-1</sup> and 15.6 kJ mol<sup>-1</sup> M<sup>-1</sup>, respectively, calculated by assuming two-state unfolding.<sup>23</sup> These  $m$ -values are typical for proteins of this size.<sup>24</sup> gpD $\Delta$ N2 and wild-type gpD are essentially indistinguishable in such equilibrium unfolding experiments (data not shown), indicating that the N-terminal region has no influence on the stability of gpD. Assuming that this lack of influence of the N-terminal region on monomer stability is true for both proteins, we suggest that the lower stability of SHP $\Delta$ N2 in comparison to gpD and gpD $\Delta$ N2 indicates that monomeric SHP (see below) may have a strongly reduced thermodynamic stability compared to monomeric gpD (13.8 kJ mol<sup>-1</sup> versus 21.6 kJ mol<sup>-1</sup>).

While SHP $\Delta$ N2 (starting at Thr23) is missing three residues whose side-chains interact with the neighboring subunits, the SHP $\Delta$ N1 variant (starting at Gly17) contains these residues. SHP $\Delta$ N1 forms trimers, albeit of lower stability (data not shown), indicating that the missing N-terminal residues contribute to constraining the crucial interacting residues Asp19, Pro20 and His22. Since we cannot prepare stable monomers of the wild-type SHP, we attempted to trap this state as an intermediate after short time refolding at low concentration (0.5  $\mu$ M), where stable trimer formation is slow (see below). While it is not possible to obtain pure monomer in this way (trimer formation already starts), we can deduce some properties of the monomer. On the one hand, it is clear from its native-like fluorescence spectrum that the monomer must be folded in a native-like structure, but since a denaturation curve similar to that of SHP $\Delta$ N2 is obtained (data not shown), it must be rather unstable. Taking these arguments together, SHP $\Delta$ N2 seems to be a valid model for monomeric SHP.

### SHP has a high kinetic stability in solution

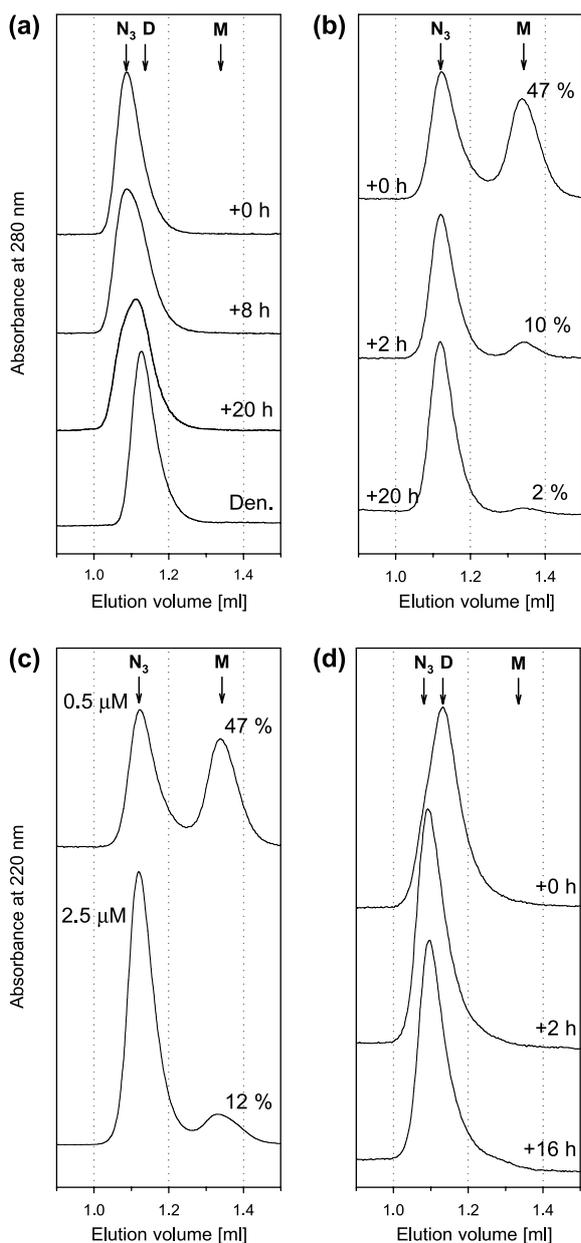
The strong hysteresis observed for equilibrium unfolding of SHP prompted us to investigate the kinetics of folding and unfolding in more detail. We thus determined SHP unfolding rates by monitoring the decrease in fluorescence intensity at 325 nm following dilution into GdmCl concentrations ranging from 3 M to 6 M, at temperatures ranging



**Figure 5.** Unfolding kinetics of trimeric SHP monitored by fluorescence emission spectroscopy. (a) Dependence of the apparent unfolding rate constants on denaturant concentration. The logarithm of the first-order rate constants is plotted as a function of GdmCl. The rate constants extrapolated to zero denaturant are  $1.3 \times 10^{-7}\text{ s}^{-1}$  (21 °C),  $3.2 \times 10^{-7}\text{ s}^{-1}$  (25.3 °C),  $7.9 \times 10^{-7}\text{ s}^{-1}$  (29.8 °C), and  $26.4 \times 10^{-7}\text{ s}^{-1}$  (34.6 °C). (b) Arrhenius plot for SHP unfolding. The logarithm of the first-order rate constants at zero denaturant is plotted as a function of temperature. The calculated Arrhenius activation energy  $E_a$  for the unfolding of SHP in the absence of denaturant is 167 kJ mol<sup>-1</sup>.

from 21 °C to 34.6 °C. The slow progress of the reaction allowed manual mixing under all conditions tested. The reactions showed only one phase and the individual unfolding rate constants were determined by fitting the reaction to a single exponential. The logarithm of the unfolding rate constants showed a linear dependence on the concentration of denaturant at all temperatures tested (Figure 5(a)). Thus, first-order rate constants of  $1.3 \times 10^{-7}\text{ s}^{-1}$  (21 °C),  $3.2 \times 10^{-7}\text{ s}^{-1}$  (25.3 °C),  $7.9 \times 10^{-7}\text{ s}^{-1}$  (29.8 °C), and  $26.4 \times 10^{-7}\text{ s}^{-1}$  (34.6 °C) for the unfolding of SHP under native conditions (zero denaturant) were estimated by extrapolation of the data. These rates are uncommonly slow and correspond to half-lives of SHP of 65, 25, 10, and 3 days, respectively.

From an Arrhenius analysis of the temperature dependence of the unfolding rates, we calculated an



**Figure 6.** Unfolding and refolding of SHP followed by gel filtration. (a) Unfolding of solution SHP shows that it is kinetically trapped as a trimer. SHP, at a concentration of  $5 \mu\text{M}$ , was unfolded in  $\text{TBS}_{150}$  (pH 7.5) containing  $2 \text{ M}$  GdmCl at  $25^\circ\text{C}$  for the times indicated and then analyzed on a Superdex 75 column using the same running buffer. A run of completely denatured SHP (Den.) is also shown. An arrow indicates the expected elution volume for a folded monomer. (b) Refolding of SHP involves a folded monomeric intermediate. Denatured SHP ( $50 \mu\text{M}$  in  $6 \text{ M}$  GdmCl) was diluted  $1:100$  into  $\text{TBS}_{150}$  (pH 7.5) and incubated for the indicated times at  $25^\circ\text{C}$  and then analyzed on a Superdex 75 column using  $\text{TBS}_{150}$  (pH 7.5) as running buffer. The percentage of the folded monomeric fraction is designated in the graph. (c) Concentration dependence of the refolding reaction. Denatured SHP ( $50 \mu\text{M}$  or  $250 \mu\text{M}$  in  $6 \text{ M}$  GdmCl) was diluted  $1:100$  into  $\text{TBS}_{150}$  (pH 7.5) and directly analyzed on a Superdex 75 column using  $\text{TBS}_{150}$  (pH 7.5) as running buffer. The percentage of the folded monomeric fraction is designated in the graph. (d) Double jump experiments. Denatured SHP ( $50 \mu\text{M}$  in  $6 \text{ M}$  GdmCl) was diluted

activation energy ( $E_a$ ) for the unfolding reaction of  $167 \text{ kJ mol}^{-1}$  in the absence of denaturant (Figure 5(b)). This energy barrier is unusually high, as typical Arrhenius activation energies for the unfolding of small proteins in the absence of denaturant range from  $60 \text{ kJ mol}^{-1}$  to  $120 \text{ kJ mol}^{-1}$ .<sup>25</sup> Another example for a kinetically very stable protein, dimeric Rop, possesses an activation energy of  $134 \text{ kJ mol}^{-1}$  even under strongly unfolding conditions of  $5 \text{ M}$  GdmCl,<sup>25</sup> which is only slightly higher than the  $128 \text{ kJ mol}^{-1}$  determined for SHP in the presence of the same amount of GdmCl. The unfolding rate constants of SHP were independent of the protein concentration over at least a 100-fold range (from  $0.05 \mu\text{M}$  to  $5 \mu\text{M}$ ), as expected for an unfolding reaction (data not shown). The observed first-order unfolding furthermore shows that no obvious intermediate is accumulating under the employed experimental conditions. Monomeric SHP $\Delta\text{N}2$ , gpD, and gpD $\Delta\text{N}2$  do not exhibit the high kinetic stability of SHP; their unfolding rate constants at  $2 \text{ M}$  GdmCl and  $25^\circ\text{C}$  are about  $10^4$  times higher than the rate for trimeric SHP (data not shown). Thus, these data provide strong evidence that the native trimeric state of SHP is the observed kinetically stable entity.

#### The kinetic stability of SHP in solution is intrinsic to its trimeric state

The unfolding of SHP was also followed by analytical gel filtration in order to elucidate if monomers of SHP could be detected as unfolding intermediates. Due to the slow kinetics of unfolding, this reaction could be followed by gel filtration, which takes less than 25 minutes from the injection to the recording. We chose to investigate the unfolding of SHP at  $2 \text{ M}$  GdmCl, because SHP is completely denatured at equilibrium at this denaturant concentration (Figure 4(b)) and because its half-life, when denaturation is started from the trimer, of 10.5 hours under these conditions (Figure 5(a)) is still much longer than the analysis time of gel filtration. The running buffer also contained  $2 \text{ M}$  GdmCl to preclude refolding of the sample during the runs. It was first determined that trimeric folded and monomeric unfolded SHP elute at  $1.09 \text{ ml}$  and  $1.13 \text{ ml}$ , respectively, under these buffer conditions (Figure 6(a)). The hydrodynamic radius of the compact trimer is thus only slightly larger than that of the flexible unfolded monomer. The time course of SHP unfolding was followed (Figure 6(a)) and it was found that the folded trimeric state directly converts to the monomeric

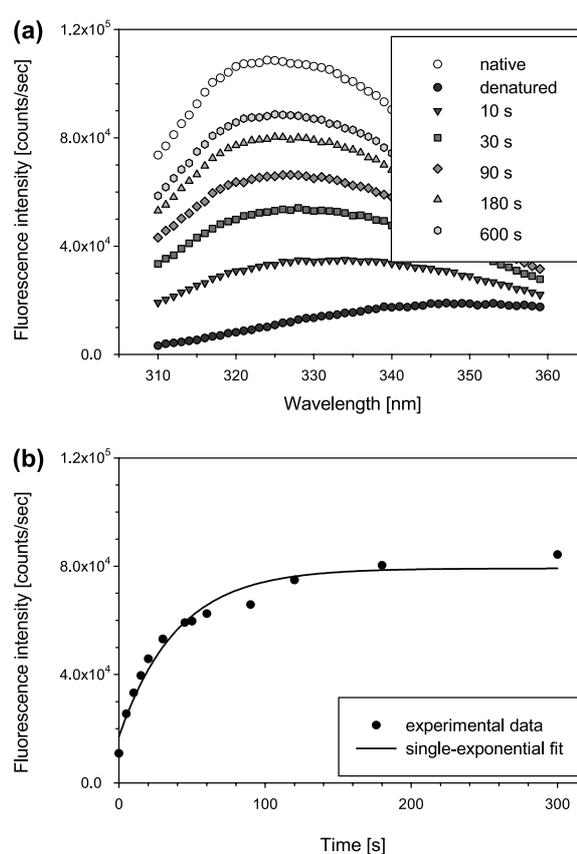
$1:100$  into  $\text{TBS}_{150}$  (pH 7.5) and incubated for the indicated times at  $25^\circ\text{C}$  and then analyzed under denaturing conditions on a Superdex 75 column using  $\text{TBS}_{150}$  (pH 7.5) containing  $2 \text{ M}$  GdmCl as running buffer. For the zero hour sample the delay between the start of the refolding and the injection was five minutes.

unfolded state without any stable monomeric intermediate detectable in 2 M GdmCl. After an incubation time of eight hours, more than about 60% of SHP is still in its trimeric state, and after 20 hours more than about 70% is denatured. We can thus conclude that the half-life of trimeric SHP in 2 M GdmCl lies between eight and 20 hours, in good agreement with 10.5 hours estimated from the fluorescence experiments (Figure 5(a)). The trimeric state of SHP must provide a barrier against unfolding, and if a folded monomeric SHP is an intermediate, it instantly denatures in 2 M GdmCl, preventing its detection. Thus, the considerable kinetic stability of SHP is due to its trimeric nature and not to an unusual stability of its subunit.

### Refolding of SHP involves a folded monomeric intermediate

Like the unfolding, the refolding of SHP was also followed by analytical gel filtration in order to elucidate if a folded monomeric intermediate is detectable during this process. Denatured SHP was diluted 1 : 100 to the final concentration of 0.5  $\mu\text{M}$  into TBS<sub>150</sub> (pH 7.5) to start refolding. It is important to note that there was a delay of five minutes between the start of the refolding and injection on the column for the zero hour sample and that this time is enough for the formation of the hydrophobic core of SHP as indicated by Trp fluorescence measurements, which indicate a native-like spectrum (data not shown). Thus, such gel filtration experiments visualize only folded species of SHP. When analyzing samples after various refolding times, we observed a SHP species eluting at 1.33 ml (Figure 6(b)). This elution volume corresponds to that of folded monomeric gpD (Figure 3). Thus, refolding of SHP involves a folded monomeric intermediate. With increasing refolding time, more and more monomers assemble into trimers. Figure 6(c) shows that the rate of trimerization observed by gel filtration is concentration-dependent, as expected for a multi-molecular reaction. These chromatograms have been obtained about five minutes after the start of refolding. This also indicates that the formation of stable trimers is relatively fast, as long as the monomer concentration is high enough (Figure 6(c)). The slow trimerization seen at monomer concentrations below 0.5  $\mu\text{M}$  (Figure 6(b)) is thus a direct result of the trimerization process that gets rate-limiting at low concentrations.

Results of a series of double jump experiments are shown in Figure 6(d). Denatured SHP was refolded at a concentration of 0.5  $\mu\text{M}$  for various times in 60 mM GdmCl (final concentration) and then applied to a gel filtration column running in 2 M GdmCl. This procedure assured that all refolded, but not yet kinetically stabilized SHP species would immediately unfold upon injection on the column. These double jump experiments show that the monomeric SHP formed during refolding (Figure 6(b)) is not stable in 2 M GdmCl.



**Figure 7.** Refolding kinetics of trimeric SHP followed by double jump fluorescence experiments. Denatured SHP (250  $\mu\text{M}$  in 6 M GdmCl) was diluted 1 : 100 into TBS<sub>150</sub> (pH 7.5) and incubated for the indicated times at 25 °C to allow refolding at the final concentration of 2.5  $\mu\text{M}$  SHP and 60 mM GdmCl. The sample was then immediately diluted 1 : 5 to denaturing conditions (TBS<sub>150</sub> (pH 7.5), 2 M GdmCl) and fluorescence emission spectra were recorded. Thereby, the formation of kinetically stable trimeric SHP can be followed. (a) Fluorescence emission spectra after various refolding times. The spectra of native and denatured SHP are also shown. Protein excitation was at 295 nm. (b) Time course of the fast phase of SHP trimerization monitored by the change in fluorescence intensity at 325 nm. The apparent first-order rate constant is  $2.5 \times 10^{-2} \text{ s}^{-1}$ .

Whereas SHP was mostly denatured in the zero hour sample, the two hour sample already consisted mainly of the kinetically stable trimer, indicated by the peak position of 1.09 ml. To determine the rate of stable trimer formation more accurately similar double jump experiments were followed by fluorescence spectroscopy (Figure 7). Denatured SHP was refolded at a concentration of 2.5  $\mu\text{M}$  for various times in 60 mM GdmCl (final concentration) at room temperature, diluted 1 : 5 into buffer containing 2 M GdmCl (final concentration) and then analyzed by fluorescence spectroscopy (Figure 7(a)). This procedure assured that all refolded, but not yet kinetically stabilized SHP species would become unfolded again before the fluorescence measurement. We thus determined the rate of formation of stable SHP by monitoring

the increase in “GdmCl-resistant” fluorescence intensity at 325 nm (Figure 7(b)). The reaction showed two phases. The fast phase reached a plateau after five minutes, and the amplitude indicates about 75% refolded trimeric SHP at this point, which is no longer able to fold in 2 M GdmCl. The residual 25% gradually convert to a GdmCl-resistant form over hours as also seen in the gel filtration experiment (Figure 6(b)). The rate constant of the first phase was determined by fitting the data to a first-order reaction, resulting in a rate constant of  $2.5 \times 10^{-2} \text{ s}^{-1}$  (at RT). When the folding of SHP was observed by fluorescence in the absence of any double jump experiment, a very similar rate constant of  $2.1 \times 10^{-2} \text{ s}^{-1}$  was obtained (data not shown). Importantly, this rate constant was identical at a protein concentration of 2.5  $\mu\text{M}$  and 0.5  $\mu\text{M}$ , and there was no further increase in fluorescence after five minutes, meaning that all SHP is folded (monomer plus trimer) at this time, but not all SHP is in the trimer state yet. We interpret the slow phase of the trimer formation (Figure 6(b)) to be a consequence of the steep concentration dependence of a trimolecular reaction. As the monomer concentration depletes, the trimer formation slows down.

Overall, these experiments show that a folded monomeric intermediate, which is not kinetically stable in 2 M GdmCl, is involved in the folding process, and the trimeric SHP is the kinetically stable species.

## Discussion

### A folding and assembly model for SHP in solution

The simplest scheme that is consistent with our folding and unfolding data collected on SHP in solution by Trp fluorescence and gel filtration is the following:



where D represents denatured monomer, which rapidly folds into a monomeric intermediate, M. M was detectable during SHP refolding followed by gel filtration (Figure 6(b)). Trp fluorescence measurements indicated that the core formation of SHP that involved burying Trp52 and Trp90 is completed after five minutes of refolding (see above). This argues that M corresponds to a form of SHP with a native-like core region. M does not possess unusually high kinetic stability as it immediately denatures in 2 M GdmCl (Figure 6(a) and (d)). However, the folding process is incomplete in the monomeric state and proceeds further to the trimer  $N_3$ , which is the native trimer in solution, whose rate of formation is concentration dependent (Figure 6(c)) as expected for a multi-molecular reaction. Unfolding experiments indicate that  $N_3$  is the kinetically stable form of SHP in solution (Figure 6(a)). The evidence for the

participation of the N terminus in stable trimer formation is that the SHP $\Delta$ N2 variant never forms stable trimers. In the monomeric SHP structure, the residues Asp19, Pro20 and His22 would be exposed to solvent, by analogy to the monomeric gpD as seen in NMR (see below). Thus, the kinetic stabilization of SHP seems to include a co-folding of the three N termini or structural rearrangements leading to the formation of the Pro/His-ring and the buried salt bridge connecting Asp19 to Lys79 of the respective neighboring subunit.

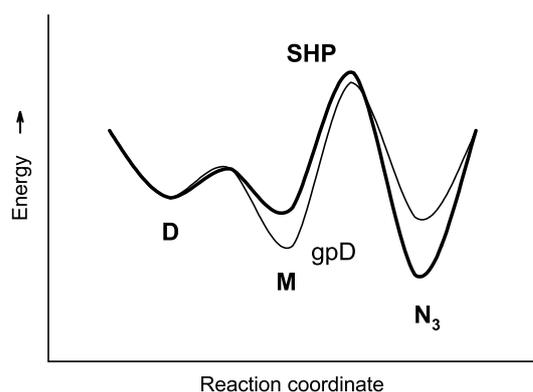
It is tempting to speculate that gpD follows the same overall folding and assembly model as SHP. In contrast to SHP, however, gpD remains as a monomer in solution (Figure 3; see below). The N-terminal residues up to His19 (corresponds to His22 of SHP) are flexible in monomeric gpD in solution, as shown by NMR experiments using  $^{15}\text{N}$  labeled gpD (H. Iwai *et al.*, unpublished results).<sup>26</sup> Thus, monomeric gpD seems to correspond to the M state of SHP. Trimeric gpD could be obtained in solution in only very small amounts, utilizing monomer concentrations above 3 mM and incubation over several weeks on ice in 40 mM Tris-Cl (pH 8.5) containing 10% (v/v) glycerol and 28% (w/v) polyethylene glycol (PEG) 3300. The gpD trimer purified by gel filtration was then stable in TBS<sub>150</sub> (pH 7.5) at low micromolar concentrations for hours, but seemed to slowly disintegrate into monomers with time (data not shown). The slow kinetics reveal the presence of a kinetic barrier between trimeric gpD and monomeric gpD, similar to that found for SHP.

Overall, these data are consistent with a multi-phasic model for SHP folding that first includes the formation of a monomer, which accumulates at low protein concentration, where trimer formation is slower, followed by monomer assembly into kinetically stable trimers, which includes the folding of the N termini into the Pro/His-ring structure and the formation of the intramolecular salt bridge employing Asp19.

### Different solution behavior of SHP and gpD

While SHP is trimeric in solution, gpD remains a stable monomer. The difference in the respective thermodynamic stabilities of the monomeric (M) and trimeric ( $N_3$ ) states of SHP and gpD determines which state is mainly present at equilibrium at a given concentration (Figure 8). For gpD, the monomeric state is the most stable state at micromolar concentrations. Purified  $N_3$  obtained after incubation at very high protein concentrations (see above) slowly dissociated into M over time. For SHP, the trimeric state  $N_3$  is the thermodynamic most stable state at micromolar concentrations. Refolding of denatured SHP (D) always results in the accumulation of  $N_3$  under these conditions.

Differences in the overall interaction energy between the respective M and  $N_3$  states make the SHP trimer more stable against dissociation than the gpD trimer. One noticeable difference in the



**Figure 8.** Schematic illustration of a possible free energy diagram for the folding of gpD and SHP in solution under physiological conditions (TBS<sub>150</sub> (pH 7.5) at room temperature) at micromolar subunit concentrations. The difference in the respective thermodynamic stabilities of the monomeric (M) and trimeric (N<sub>3</sub>) states determines which of them is mainly present at equilibrium. For gpD, the monomeric state M is the thermodynamic most stable state at micromolar concentrations (thin line). Purified N<sub>3</sub> obtained after incubation at very high protein concentrations (see the text) slowly dissociated into M over time. For SHP, the trimeric state N<sub>3</sub> is the thermodynamic most stable state at micromolar concentrations (thick line). Refolding of denatured SHP (D) always results in the accumulation of N<sub>3</sub> under these conditions (Figure 6(b)). M of SHP appears to be less stable with respect to D than M of gpD, as deduced from the behavior of SHPΔN2 (see the text).

appearance of the trimers in the crystal structures is the number of intersubunit H-bonds in SHP (21 in total), compared to only nine in gpD. Another difference between SHP and gpD affects their M states. Assuming that SHPΔN2 is a valid model for the M state of SHP (see Results for a summary of the arguments), it follows that the M state of SHP has a significantly reduced thermodynamic stability compared to monomeric gpD (13.8 kJ mol<sup>-1</sup> versus 21.6 kJ mol<sup>-1</sup>). Such differences and the steep concentration dependence inherent in a monomer/trimer equilibrium may explain why the gpD trimer formation can only occur at the extremely high concentrations during crystal growth, while SHP is a stable trimer at micromolar concentrations. We have summarized these findings in an energy diagram in Figure 8.

### Kinetic stabilization by protein oligomerization

Trimeric SHP possesses a high kinetic stability in solution with a half-life under physiological conditions (zero denaturant) of 25 days (Figure 5(a)) and an Arrhenius activation energy for unfolding of 167 kJ mol<sup>-1</sup>. Furthermore, we demonstrated that the trimeric form of SHP is the kinetically stable entity (Figure 6(a)). When the partly disordered N terminus of SHP is removed up to His22, the resulting SHPΔN2 is monomeric, shows similar equilibrium unfolding as gpD and gpDΔN2, and

equilibrates as fast as is typical for small proteins. It follows that the trimeric state of SHP is directly linked to an extraordinary kinetic stability. This stability is seen by the fact that the protein still runs as a trimer in SDS-containing gels, and that an unfolding rate is measured, which is almost 10<sup>4</sup>-fold slower than the rate of gpD, gpDΔN2 and SHPΔN2. Since the trimeric state is lost upon removal of the N-terminal residues in SHPΔN2, it is in this region that important components of the kinetic stabilization must reside.

There are different structural features that may contribute to this high kinetic stability of trimeric SHP. First, the fact that SHP is trimeric leads, *per se*, to a high-energy barrier for dissociation, since for the first subunit to dissociate from the trimer, two subunit interactions must be broken simultaneously. The higher number of intersubunit H-bonds in SHP compared to gpD may contribute to its greater stability. If the trimer is associating as to engage structural elements in the interface which are crucial for the initiation of denaturation, this would also kinetically stabilize the trimer toward denaturation. Nevertheless, this alone cannot be responsible for the extraordinary stability, as SHPΔN2, which has almost all the residues of the interface, is extremely unstable, and even SHPΔN1, which carries the whole interface region, is still significantly less stable than wt SHP (see Results).

A second element must thus be invoked, which involves the structure at the N terminus. Two features specifically contributed by this region are the Pro/His ring (residues 20 and 22), as well as a salt bridge from Asp19 to Lys79 of the neighboring subunit. In order to fix these residues in their proper orientation, the N-terminal stretch (visible in the electron density from Val12 on) must be pinned down to the bulk of the domain. Two main chain H-bonds (between Ala16 N and Thr82 O, and between Ile 14 O and Ala84 N), as well as several hydrophobic contacts (see Results) stabilize these interactions, whereas this region is disordered in the structure of trimeric gpD. Thus, it becomes clear why the SHPΔN1 variant, in which all residues are present, which interact with the other subunits, dissociates and denatures significantly faster than the wild-type, as Asp19, Pro20 and His22 are not conformationally as constrained in this deletion mutant. The buried intermolecular salt bridge between Asp19 and Lys79 of the neighboring subunit is also difficult to break in the presence of all the other interactions, which constrain any fluctuations. Beside this interaction with Asp19, Lys79 also interacts with the backbone oxygen atoms of Ala107 and Ser109 from the neighboring subunit and thereby locks the N and C termini of its neighboring subunit in place (Figure 2). Furthermore, its own main-chain NH and that of its neighbor Ser80 fix the carbonyl group of His22 of its own subunit (Figure 1(b)). Additionally, the breaking of the buried salt bridge between Asp19 and Lys79 will not only be characterized by a high kinetic barrier<sup>27</sup> because of these steric effects, but

also because of the energetic cost of charge separation. The high cost of breaking of this charged interaction is only compensated at the end of the reaction when solvation of these side-chains can occur.

It is interesting to note that proline-mediated arm exchange is a commonly used oligomerization mechanism.<sup>28</sup> Such a role of prolines is probably especially important when the arm exchange occurs in the vicinity of a symmetry axis, as it is the case for SHP.

### Kinetic stability of viruses

Virus particles are released into the environment after completion of their assembly. They must be able to resist harsh environmental conditions until they find a new host to infect. Their capsid must also not fall apart at essentially infinite dilution. While many viruses must also be able to disassemble their capsids upon infection, icosahedral bacteriophages with tails can “inject” the DNA from a stable capsid. Their capsid assembly can thus be quasi-irreversible. Virus capsid assembly approaches equilibrium in biologically useful time, whereas dissociation does not.<sup>15</sup> This hysteresis means that assembled capsids will behave as though they are considerably more stable than they are thermodynamically; they possess a high kinetic stability. The multimeric nature of virus capsids, *per se*, already makes the disassembly, which requires concerted motions of the peptide chain, very difficult. In order for the first subunit to dissociate from the capsid, several subunit interactions must be broken simultaneously. This is a very difficult task, especially in view of the often-found mutually intertwined terminal peptides of viral capsid proteins. The distinction between assembly and disassembly of particles is further enhanced in some viruses by a maturation step such as proteolysis, crosslinking (disulfide bonds or isopeptide linkages) or, frequently, conformational change.<sup>15,29–32</sup>

Besides these strategies, some icosahedral phages, such as phage  $\lambda$  and its homolog phage 21, further stabilize their capsids by using a clamping protein at the 3-fold sites. This is the function of gpD and SHP, and the T4 protein Soc plays a similar role in phage T4.<sup>33</sup> Clearly, such a protein must also be stable by itself, and we show here that by trimerization of SHP, a 10,000-fold decrease in unfolding rate is achieved, compared to the monomeric state. The protein has apparently found an association mode, which, for all practical purposes, blocks all pathways for unfolding. It is reasonable to assume that SHP would even possess a higher kinetic stability while bound to the phage capsid, as the underlying capsid surface interacts with the bottom side of the trimer and with the N-terminal residues which are flexible in solution.<sup>26</sup> Overall, SHP seems to stabilize the phage capsid, at least partly, through its intrinsic high kinetic stability.

### Implications

This comprehensive analysis of SHP has highlighted a few remarkable features of proteins, which may be of general importance. First, a kinetic stabilization against denaturation successfully addresses the biologically relevant demand for stability. Functional stability does not necessarily have to be reflected by an extreme equilibrium thermodynamic stability. Second, the tight and interwoven interactions between the neighboring subunits may be the decisive feature for kinetic stability, and this interaction apparently makes the coherent motions involved in denaturation almost inaccessible to the protein.

SHP appears to constitute a model for a protein, which very successfully avoids the major pathways for denaturation for a maximal duration. It may be worthwhile to analyze in detail where this extreme difference between the kinetic stability of the monomeric and trimeric state comes from, and how such features can be engineered into other proteins. Clearly, not all oligomeric proteins are as stable, and thus the exact mode of trimerization must be part of the solution. In conclusion, the extreme selection pressure of phages for capsid stability may help unravel features of proteins directly useful in biotechnology.

### Materials and Methods

#### Expression and purification of SHP and gpD variants

SHP (114 amino acid residue, 11.8 kDa), SHP $\Delta$ N1 (GT-SHP(G17-Pro114)) and gpD (109 amino acid residue, 11.4 kDa) were expressed in soluble form in the cytoplasm of the *Escherichia coli* strain BL21(DE3)[pLysS] (Stratagene) using the T7 promoter-based bacterial expression plasmids pAT122, pAT179 and pAT101, respectively, and purified to near homogeneity as described.<sup>19</sup> The gene for SHP originates from the bacteriophage  $\lambda$ -21 hybrid 19 (kindly provided by M. Feiss).<sup>34</sup> The expression plasmids pAT147 and pAT143, encoding His-tagged SHP $\Delta$ N2 (MRGSH<sub>6</sub>GS-SHP(T23-P114)) and gpD $\Delta$ N2 (MRGSH<sub>6</sub>-GSGSMG-gpD(T20-V109)), respectively, were constructed by PCR amplification of the appropriate fragments from pAT122 and pAT101 and subcloned into a pQE30 (Qiagen) derived vector. SHP $\Delta$ N2 and gpD $\Delta$ N2 were expressed in the cytoplasm of the *E. coli* strain XL1-blue (Stratagene) in soluble form and purified to near homogeneity by Ni<sup>2+</sup>-chelate chromatography using reagents and protocols from Qiagen. All proteins are very well expressed (>200 mg/l in standard shake flask cultures), highly soluble and are stable on storage. The protein concentrations were determined as described by Gill & von Hippel.<sup>35</sup> All protein concentrations throughout this paper refer to the concentrations of monomers or subunits.

#### Crystallization of SHP

Lyophilized SHP protein was dissolved in 25 mM Tris buffer (pH 7.5), and concentrated to 30 mg/ml. Crystallization was performed by the hanging-drop vapor diffusion method at 22 °C. Crystal screen I (Hampton Research) was used for the initial screening. A small,

plate-shaped crystal was found from screen no. 20. After further refinement, crystals could be grown under two related crystallization conditions. One of them was 100 mM sodium acetate (pH 4.5), 22–25% (w/v) PEG-MME 2000, 200 mM MgSO<sub>4</sub>, 70–100 mM glycine (form I crystals). The other condition was 100 mM sodium acetate (pH 4.5), 26–27% PEG-MME 5000, 100 mM MgCl<sub>2</sub> (form II crystals). Form I crystals would grow to the size of 0.3–0.7 mm in ~20 days. They were shaped as hexagonal plates and were usually stacked. Form II crystals could grow to a size of around 0.3 mm in ~40 days. These crystals were shaped as cubes and would grow from the precipitate. For data collection at 100 K, crystals of either form were soaked in mineral oil for ~40 seconds and then flash-frozen in a nitrogen stream.

### Data collection for SHP

Form I crystals of SHP belong to the monoclinic space group C2 with the unit cell parameters  $a=100.1$  Å,  $b=57.7$  Å,  $c=62.0$  Å,  $\beta=117.1^\circ$ ,  $V_M=2.21$  Å<sup>3</sup>/Da. Three SHP molecules are present in each asymmetric unit. A data set extending to 2.37 Å was collected using a MAR345 image plate detector (MAR Research, Hamburg, Germany) mounted on a Rigaku rotating anode generator (CuK $\alpha$  radiation, 1.5418 Å) at 100 K. These crystals exhibited very high mosaicity (1.2–1.9°), but data could still be processed. Form II crystals are rhombohedral, space group R3, with the hexagonal setting unit cell parameters  $a=b=53.9$  Å,  $c=77.5$  Å,  $V_M=1.81$  Å<sup>3</sup>/Da. Only a single SHP molecule is present in the asymmetric unit of this crystal form. A data set extending to 1.5 Å was collected at 100 K using an ADSC Quantum 4 CCD detector on the synchrotron beamline X9B at the National Synchrotron Light Source, Brookhaven National Laboratory, Upton, New York. Data for both crystal forms were integrated and scaled using the HKL2000 program suite.<sup>36</sup> Data collection statistics are summarized in Table 1.

The structure of SHP was solved by molecular replacement with the program AMoRe.<sup>37</sup> A search model for the crystal form I was chain A of gpD (Protein Data Bank accession code 1c5e). The fully automated script was run in the resolution range 15–3.5 Å, yielding the positions of three molecules, with the correlation coefficient of 0.453, and the  $R$ -factor of 43.6%. Form II was solved using the coordinates of chain A of form I and the same resolution limits of the data, with a correlation coefficient of 0.587 and  $R$ -factor of 38.3%.

The model for SHP in form I crystals was refined using CNS 1.0 at the resolution range of 25.0–2.37 Å.<sup>38</sup> The non-crystallographic symmetry (NCS) restraints with an energy barrier of 300 kcal mol<sup>-1</sup> Å<sup>-2</sup> between the three subunits were maintained throughout the refinement. At the first stage of refinement, amino acid residues different from those present in gpD were mutated to those corresponding to the sequence of SHP. The model was rebuilt with the program O using both  $3F_o - 2F_c$  and  $F_o - F_c$  maps.<sup>39</sup> In addition to protein atoms, 387 water molecules have been added to the model. The  $R$ -value for all reflections (45.0–2.37 Å) is 21.0% ( $R_{free}=24.3\%$ ). The model for form II was refined using SHELXL at the resolution range of 40.0–1.5 Å. The model was rebuilt with the program O using both  $2F_o - F_c$  and  $F_o - F_c$  maps. Three full cycles of remodeling and refinement were performed, with the refinement of individual anisotropic  $B$ -factors for all atoms introduced in the last cycle. In addition to protein atoms, 167 water molecules have been added to the model. The  $R$ -value for all reflections in the 40.0–1.5 Å range is 12.6% ( $R_{free}$  of 18.5%).

The geometrical properties of the model were assessed with the program PROCHECK<sup>40</sup> and the secondary structure elements were assigned by the program PROMOTIF.<sup>41</sup> The surface charge potential was calculated by GRASP,<sup>42</sup> and this program was also used to generate surface displays. Other Figures were prepared with MOLSCRIPT<sup>43</sup> or BOBSCRIPT<sup>44</sup> and rendered with Raster3D.<sup>45</sup>

The conformation of Ser93 in SHP is very unusual, since that amino acid appears to be D rather than L-serine. That interpretation is not obvious in the lower-resolution and poorer quality structure obtained from crystal form I, but is much more certain in the maps obtained for form II. In the latter case, when Ser93 was refined in the usual L-configuration, its side-chain was not well positioned in the map and could not be made to fit the density, which was otherwise excellent for this stretch of the polypeptide chain. The torsion angle  $\omega$  for this residue was refining to over 30° away from planarity in the L-conformation. To remove any possible model bias, this residue was mutated to a glycine and the refinement was continued. After the completion of this round of refinement, the map showed that the position of C $\beta$  was different from the expected configuration. Two models, one including L-Ser and the other D-Ser, were refined independently and compared with the map generated from the glycine model or with a map calculated from a refinement cycle in which residues 92–94 were completely removed. In both cases, a model including D-serine fits much better than the one with L-serine and, in addition, the stereochemistry is much superior in the former case than in the latter.

### Size-exclusion chromatography (gel filtration)

All gel filtration experiments were done on a HPLC system (SMART system; Amersham Pharmacia Biotech) using a Superdex 75 PC 3.2/30 column (Amersham Pharmacia Biotech) at 25 °C. This column has a bed volume of 2.4 ml, a void volume of 0.89 ml (experimentally determined using blue dextran) and an optimal separation range from 3 kDa to 70 kDa. All runs were done at 60  $\mu$ l min<sup>-1</sup> in TBS<sub>150</sub> (20 mM Tris, 150 mM NaCl (pH 7.5)) containing variable amounts of guanidinium chloride (GdmCl) as indicated. The sample volume was 50  $\mu$ l. Cytochrome *c* (12.4 kDa), carbonic anhydrase (29 kDa) and bovine serum albumin (66 kDa) were used as standard proteins to obtain a calibration curve from which the apparent molecular masses of the sample proteins could be calculated.

### Determination of equilibrium unfolding curves

SHP, SHP $\Delta$ N2 and gpD samples for fluorescence measurements were prepared in TBS<sub>150</sub> (pH 7.5) containing 0.5  $\mu$ M protein and variable amounts of GdmCl. Denaturant concentrations were determined refractometrically.<sup>23</sup> After incubation for ten days (SHP) or 12 hours (SHP $\Delta$ N2, gpD) at 25 °C, fluorescence emission spectra were recorded at 25 °C with a PTI Alpha Scan spectrofluorimeter (Photon Technologies Inc.). Three fluorescence spectra per sample were averaged. The protein was excited at 295 nm and the emission spectra were recorded from 310 nm to 360 nm. For the equilibrium unfolding curves, the fluorescence intensity was followed at 325 nm and the data were fit to a two-state unfolding model<sup>23</sup> using the program SigmaPlot (SPSS Inc.). Thereby,  $[D]_{1/2}$ ,  $\Delta G$ , and  $m$ -values were extracted. SHP does not follow such a two-state model and thus  $\Delta G$  and  $m$ -values could not be determined for this protein.

### Kinetics of protein unfolding

Unfolding rates were measured at the various temperatures (from 21 °C to 34.6 °C) using a PTI Alpha Scan spectrofluorimeter. Protein stock solutions in TBS<sub>150</sub> (pH 7.5) were diluted 1 : 20 (v/v) by manual mixing into TBS<sub>150</sub> (pH 7.5) containing the appropriate amounts of GdmCl. The final protein concentration was 0.5 μM. Final denaturant concentrations were determined refractometrically.<sup>23</sup> Protein excitation was at 295 nm and the decrease in fluorescence emission intensity was recorded at 325 nm. The kinetic traces were fit to single exponentials. Unfolding rates in the absence of denaturant were estimated by linear extrapolation using the equation  $\ln(k_u(\text{obs}, D)) = \ln(k_u(\text{H}_2\text{O})) + m_u[D]/RT$ ,<sup>46</sup> where  $k_u(\text{obs}, D)$  is the apparent first-order rate constant for unfolding at denaturant concentration  $[D]$ ,  $k_u(\text{H}_2\text{O})$  is the apparent rate constant in the absence of denaturant, and  $m_u$  indicates the dependence on denaturant, which is a measure of the change in solvent accessibility of the unfolded state,  $R$  is the gas constant and  $T$  is the absolute temperature. Arrhenius activation energies for the unfolding of SHP at a certain denaturant concentration were determined using the Arrhenius equation  $k_u(\text{obs}, T) = A \exp(-E_a/RT)$ , where  $k_u(\text{obs}, T)$  is the apparent first-order rate constant for unfolding at the temperature  $T$ ,  $A$  is a temperature-independent parameter,  $E_a$  is the activation energy,  $R$  is the gas constant and  $T$  is the absolute temperature. All data fitting was done with SigmaPlot (SPSS Inc.).

### Kinetics of SHP refolding

The refolding rate of SHP was measured at 25 °C by double jump experiments. Denatured SHP (250 μM in 6 M GdmCl) was diluted 1 : 100 into TBS<sub>150</sub> (pH 7.5) and incubated for variable times at 25 °C to allow refolding at 2.5 μM. The sample was then immediately diluted 1 : 5 to denaturing conditions (TBS<sub>150</sub> (pH 7.5), 2 M GdmCl) and the fluorescence emission spectrum was recorded using a PTI Alpha Scan spectrofluorimeter (Photon Technologies Inc.). The protein was excited at 295 nm and the emission spectra were recorded from 310 nm to 360 nm. Three fluorescence spectra per sample were averaged. To follow the time course of SHP refolding the fluorescence intensity at 325 nm was extracted from the average spectra and the data were fit to a single exponential using the program SigmaPlot (SPSS Inc.).

### Protein Data Bank accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank, with accession numbers 1TD3 and 1TD4 for the structures of forms I and II, respectively.

### Acknowledgements

We thank H. Iwai for stimulating discussions and Z. Dauter for help in data collection on beamline X9B at NSLS, Brookhaven National Laboratory. This work was funded in part by the Schweizerische Nationalfonds (grant 3100-065344).

### References

- Baker, D. & Agard, D. A. (1994). Kinetics versus thermodynamics in protein folding. *Biochemistry*, **33**, 7505–7509.
- Cunningham, E. L., Jaswal, S. S., Sohl, J. L. & Agard, D. A. (1999). Kinetic stability as a mechanism for protease longevity. *Proc. Natl Acad. Sci. USA*, **96**, 11008–11014.
- Plaza del Pino, I. M., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. (2000). Lower kinetic limit to protein thermal stability: a proposal regarding protein stability *in vivo* and its relation with misfolding diseases. *Proteins: Struct. Funct. Genet.* **40**, 58–70.
- Fersht, A. R. (1999). Protein stability. In *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding* (Fersht, A. R., ed), pp. 508–539, Freeman, New York.
- Taverna, D. M. & Goldstein, R. A. (2002). Why are proteins marginally stable? *Proteins: Struct. Funct. Genet.* **46**, 105–109.
- Sinclair, J. F., Ziegler, M. M. & Baldwin, T. O. (1994). Kinetic partitioning during protein folding yields multiple native states. *Nature Struct. Biol.* **1**, 320–326.
- Chen, J., Wharton, S. A., Weissenhorn, W., Calder, L. J., Hughson, F. M., Skehel, J. J. & Wiley, D. C. (1995). A soluble domain of the membrane-anchoring chain of influenza virus hemagglutinin (HA2) folds in *Escherichia coli* into the low-pH-induced conformation. *Proc. Natl Acad. Sci. USA*, **92**, 12205–12209.
- Lai, Z., McCulloch, J., Lashuel, H. A. & Kelly, J. W. (1997). Guanidine hydrochloride-induced denaturation and refolding of transthyretin exhibits a marked hysteresis: equilibria with high kinetic barriers. *Biochemistry*, **36**, 10230–10239.
- Baskakov, I. V., Legname, G., Prusiner, S. B. & Cohen, F. E. (2001). Folding of prion protein to its native alpha-helical conformation is under kinetic control. *J. Biol. Chem.* **276**, 19687–19690.
- Barrientos, L. G., Louis, J. M., Botos, I., Mori, T., Han, Z., O'Keefe, B. R. *et al.* (2002). The domain-swapped dimer of cyanovirin-N is in a metastable folded state. Reconciliation of X-ray and NMR structures. *Structure*, **10**, 673–686.
- Fasshauer, D., Antonin, W., Subramaniam, V. & Jahn, R. (2002). SNARE assembly and disassembly exhibit a pronounced hysteresis. *Nature Struct. Biol.* **9**, 144–151.
- Sohl, J. L., Jaswal, S. S. & Agard, D. A. (1998). Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature*, **395**, 817–819.
- Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. (2002). Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature*, **415**, 343–346.
- Smith, M. P. & Feiss, M. (1993). Sequence analysis of the phage 21 genes for prohead assembly and head completion. *Gene*, **126**, 1–7.
- Zlotnick, A. & Stray, S. J. (2003). How does your virus grow? Understanding and interfering with virus assembly. *Trends Biotechnol.* **21**, 536–542.
- Imber, R., Tsugita, A., Wurtz, M. & Hohn, T. (1980). Outer surface protein of bacteriophage lambda. *J. Mol. Biol.* **139**, 277–295.
- Castagnoli, L., Zucconi, A., Quondam, M., Rossi, M., Vaccaro, P., Panni, S. *et al.* (2001). Alternative bacteriophage display systems. *Comb. Chem. High Throughput Screening*, **4**, 121–133.

18. Dokland, T. & Murialdo, H. (1993). Structural transitions during maturation of bacteriophage lambda capsids. *J. Mol. Biol.* **233**, 682–694.
19. Yang, F., Forrer, P., Dauter, Z., Conway, J. F., Cheng, N., Cerritelli, M. E. *et al.* (2000). Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nature Struct. Biol.* **7**, 230–237.
20. Forrer, P. & Jaussi, R. (1998). High-level expression of soluble heterologous proteins in the cytoplasm of *Escherichia coli* by fusion to the bacteriophage Lambda head protein D. *Gene*, **224**, 45–52.
21. Bailey, D., Cooper, J. B., Veerapandian, B., Blundell, T. L., Atrash, B., Jones, D. M. & Szelke, M. (1993). X-ray-crystallographic studies of complexes of pepstatin A and a statine-containing human renin inhibitor with endothiapsin. *Biochem. J.* **289**, 363–371.
22. Di Marco, S. & Priestle, J. P. (1997). Structure of the complex of leech-derived trypsin inhibitor (LDTI) with trypsin and modeling of the LDTI-trypsin system. *Structure*, **5**, 1465–1474.
23. Pace, C. N. & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure: a Practical Approach* (Creighton, T. E., ed), pp. 300–321, IRL Press, Oxford.
24. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
25. Munson, M., Anderson, K. S. & Regan, L. (1997). Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude. *Fold. Des.* **2**, 77–87.
26. Iwai, H., Forrer, P., Plückthun, A. & Guntert, P. (2004). Assignments of  $^1\text{H}$  and  $^{15}\text{N}$  resonances of the bacteriophage lambda capsid stabilizing protein gpD. *J. Biomol. NMR*, **28**, 89–90.
27. Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA*, **93**, 2629–2634.
28. Bergdoll, M., Remy, M. H., Cagnon, C., Masson, J. M. & Dumas, P. (1997). Proline-dependent oligomerization with arm exchange. *Structure*, **5**, 391–401.
29. Dokland, T. (2000). Freedom and restraint: themes in virus capsid assembly. *Struct. Fold. Des.* **8**, R157–R162.
30. Namba, K. (2001). Roles of partly unfolded conformations in macromolecular self-assembly. *Genes Cells*, **6**, 1–12.
31. Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W. & Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science*, **289**, 2129–2133.
32. Golmohammadi, R., Fridborg, K., Bundule, M., Valegard, K. & Liljas, L. (1996). The crystal structure of bacteriophage Q beta at 3.5 Å resolution. *Structure*, **4**, 543–554.
33. Steven, A. C., Greenstone, H. L., Booy, F. P., Black, L. W. & Ross, P. D. (1992). Conformational changes of a viral capsid protein. Thermodynamic rationale for proteolytic regulation of bacteriophage T4 capsid expansion, co-operativity, and super-stabilization by soc binding. *J. Mol. Biol.* **228**, 870–884.
34. Feiss, M., Fisher, R. A., Siegele, D. A., Nichols, B. P. & Donelson, J. E. (1979). Packaging of the bacteriophage lambda chromosome: a role for base sequences outside cos. *Virology*, **92**, 56–67.
35. Gill, S. C. & Von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
36. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
37. Navaza, J. (1994). An automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
39. Jones, T. A. & Kjeldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208.
40. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
41. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.
42. Nicholls, A. (1992). *GRASP: Graphical Representation and Analysis of Surface Properties*. Columbia University, New York, NY.
43. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
44. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallog. sect. D*, **55**, 938–940.
45. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.
46. Eftink, M. R. & Shastry, M. C. (1997). Fluorescence methods for studying kinetics of protein-folding reactions. *Methods Enzymol.* **278**, 258–286.
47. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.

Edited by F. Schmid

(Received 28 April 2004; received in revised form 9 September 2004; accepted 15 September 2004)